DOCUMENT RESUME

ED 400 333 TM 025 738

AUTHOR Bridgeman, Brent; And Others
TITLE The Reliability of Document-Based Essay Questions on Advanced Placement History Examinations.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-96-5
PUB DATE Feb 96
NOTE 31p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Advanced Placement; *Constructed Response; Correlation; *Essay Tests; European History; High Schools; *High School Students; *Reliability; *Scores; United States History
IDENTIFIERS *Advanced Placement Examinations (CEEB)

ABSTRACT

If a student writes two essays, the score reliability can be estimated from the correlation between essays. However, if the essays are in different modes or require different skills, the reliability may be underestimated from the correlation. In Advanced Placement history examinations, students wrote one standard essay and one essay that required analysis and synthesis of historical documents that were included with the question statement. If these Document-Based Questions (DBQs) were assessing substantially different skills from the standard essays, then the reliability of DBQ scores would be underestimated from their correlation with a standard essay score. A sample of 1,045 U.S. history students and 891 European history students participated in a special study in which they wrote essays for either 2 DBQ questions and 1 standard essay question or 3 standard essay questions. The DBQ correlated as highly with a standard essay as with another DBQ, suggesting that the simple correlation of the two types of scores did not underestimate the reliability of the essay scores. Appendix A lists standard essay topics, and Appendix B presents DBQ topics. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-96-5

ED 400 333

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

THE RELIABILITY OF DOCUMENT-BASED ESSAY QUESTIONS ON ADVANCED PLACEMENT HISTORY EXAMINATIONS

Brent Bridgeman
Rick Morgan
Ming-mei Wang

Educational Testing Service
Princeton, New Jersey
February 1996

BEST COPY AVAILABLE

TM025738

The Reliability of Document-Based Essay Questions on Advanced Placement History Examinations

Brent Bridgeman
Rick Morgan
and
Ming-mei Wang

January, 1996

Copyright © 1996. Educational Testing Service. All rights reserved.

Abstract

If a student writes two essays, the score reliability can be estimated from the correlation between the essays. However, if the essays are in different modes or require different skills, the reliability may be underestimated from the correlation. In Advanced Placement history examinations, students wrote one standard essay and one essay that required analysis and synthesis of historical documents that were included with the question statement. If these document-based questions (DBQ) were assessing substantially different skills from the standard essays, then the reliability of the DBQ scores would be underestimated from their correlation with a standard essay score. A sample of 1045 U. S. history students and 891 European history students participated in a special study in which they wrote essays for either 2 DBQ questions and one standard essay question or 3 standard essay questions. The DBQ correlated as highly with a standard essay as with another DBQ, suggesting that the simple correlation of the two types of scores did not underestimate the reliability of the essay scores.

The Reliability of Document-Based Essay Questions on Advanced Placement History Examinations

Methods for evaluating score reliability (as distinct from rater reliability) require that each student respond to at least two different questions that are assumed to be parallel measures of the same underlying construct. With multiple-choice tests this is typically not a problem; each student takes dozens of questions. Reliability may be estimated from the question intercorrelations or summed item variances compared to total test variance (e.g., KR-20 or coefficient alpha). Although these internal consistency measures can be artifically inflated by dependencies among items (as when several items are based on a single reading passage) or by test speededness, alternative methods are available such as investigating correlations among separately-timed groups of parallel items. These techniques are of little value in essay examinations when an examinee may write only one essay, or may write more than one essay but the essays are not considered to be parallel. Such a situation arises in the Advanced Placement (AP) examinations in U.S. and European History.

Each AP history examination consists of a multiple-choice section and two essays.[1] As described in the published course description in U.S. History, multiple choice questions "are designed to test the students' factual knowledge, breadth of preparation, and knowledge-based analytical skills. Essay questions are designed, additionally, to make it possible for students from widely differing courses to demonstrate their mastery of historical interpretation and their ability to express their views and knowledge in writing" (College Board, 1993, p. 19). Thus, the multiple-choice and essay sections should be related, but they are intended to measure somewhat

[1]In 1994, a third essay was added; this analysis is based on the examination as it was in 1993. Conclusions should generalize to the current examination.

different underlying constructs and are clearly not parallel measures. The two essay questions also are not designed to be strictly parallel. One essay question type, referred to as the standard essay question, poses a question in a sentence or two, and students have a suggested time of 50 minutes (for U.S. History) or 45 minutes (for European History) to compose an answer. (The total time limit for writing both essays is enforced, but the division of time between the two essays is merely suggested. Time limits and other details given here are accurate for the 1993 examinations that were used in this research; as a constantly evolving program, these details may change from year to year.) The answers to the essays are judged holistically "on the strength of the thesis developed, the quality of the historical argument, and the relevance of examples rather than coverage per se. Others will be answered equally well by considering representative events, interests, or individuals in depth" (p. 19-20). Although the AP Program provides broad outlines of content coverage to participating teachers, considerable local autonomy is exercised and some programs may emphasize certain issues or periods more than others. Accordingly, students are given a choice of 5 or 6 topics so that they can display their knowledge and analytical skills on a topic that they know well. Typical standard essay questions are presented in Appendix A.

The other type of essay question is referred to as the Document-Based Question (DBQ). In the DBQ the student is given 10 to 20 short documents to study and then asked to use the documents to support the answer to a question. The documents usually include paragraph-length quotations from original sources and may also include such sources as maps, graphs, pictures, and political cartoons. The student has 15 minutes to read the documents and a suggested time of 40 minutes (for U.S. History) or 45 minutes (for European History) to compose the answer. The official description of the U.S. History DBQ notes, "The required DBQ differs from the standard

essays in its emphasis on the ability to analyze and synthesize historical data and assess verbal, quantitative, or pictorial materials as historical evidence"(p. 20). Although the DBQ formats in U.S. and European History have much in common, a critical difference is that the U.S. examination requires that outside knowledge of particular events or trends "be incorporated into the students' essay if the highest scores are to be earned" (p. 21) while the European examination does not require incorporation of any outside knowledge into the response. The description of the European History DBQ states, "The primary purpose of the document-based essay question is not to test students' prior knowledge of subject matter, but rather to evaluate their ability to formulate and support an answer from documentary evidence" (p. 64). The DBQs from the 1993 examinations are presented in Appendix B.

Because each student writes only one standard essay and one DBQ, there is no way to evaluate the score reliability of either essay. Current program practice estimates the reliability of the essay score from the correlation between the scores on the two types of essays. This method is used not because it is believed to be ideal, but because there is no alternative. If each type of essay question is evaluating a somewhat different skill, and if students differ on these skills, then the correlation between the two essays could underestimate the reliability.

Previous research suggests that the underestimation may be trivial because differences among essay types (or modes) are small relative to differences among topics. In a comprehensive study of English essays, college students wrote two essays in each of three modes (narrative, expository, and persuasive) over the course of a semester (Breland, Camp, Jones, Morris, & Rock, 1987). Confirmatory factor analysis indicated that the greatest improvement over a single-factor model occured when topic factors (not mode factors) were fitted. Nevertheless, there was

some evidence that the persuasive essays were measuring some unique variance, but the administrative conditions for these essays were also quite different. The persuasive essays were assigned as extended out-of-class exercises while essays in the other modes were written in class under time restrictions. When the mode scores and individual topic scores were correlated with outside variables (such as scores on multipe-choice tests and grades), there was little evidence for discriminant validity of the mode scores. The authors concluded, "there appears to be more differentiation with respect to topic within mode than there is among modes" (p. 45).

Carlson, Bridgeman, Camp, and Waanders (1985) studied essays written in English by foreign college students. Each student wrote two essays on compare/contrast topics and two essays on descriptions of charts or graphs. Correlations across modes were as high as correlations between topics within modes. However, the underlying task in both modes may have been quite similar dispite differences in the outward appearance of the questions. The chart and graph topics implicitly required students to compare and contrast the visual stimuli that were presented. The modes differed with respect to the presence or absence of these visual stimuli but otherwise may have required very similar analytical skills and writing abilities. A study by Quellmalz, Capell, and Chou (1982) argued for the potential importance of discourse mode effects, but their results are difficult to evaluate in the current context because each student wrote only one essay in each mode.

Since 1991 the free response section of the AP Spanish Literature examination has required examinees to write three essays. Two of the essays require the examinee to respond to prose passages while the third essay requires a response to a poetry passage. The range of correlations (since 1991) between the two prose scores is .48 to .51 while correlations of the

prose scores with the poetry scores range from .34 to .40. This suggests a mode effect, although a separate estimate of the reliability of essays written in response to a poetry prompt is not possible because each student writes only a single essay based on a poetry prompt. Thus, the poetry prompt may be reliably measuring a different construct, or it could simply be assessing the same construct as the prose prompts but less reliably.

For the AP Art History examination, examinees write essays based on two different types of stumuli. In the first part of the free response section, essays are written in response to each of seven sets of one or two slides. In the second section, two long essays are written in response to questions which address important problems in the history of art. Since 1991, the average correlation of the total score on the first section with the second section is .59. Given the first section reliability of .70 and the second section reliability of .54, it appears that mode effects are small.

For the AP history examinations, the extent to which the standard essay and DBQ measure the same dimension is unknown. The current study evaluated the reasonableness of using the correlation between these two scores to estimate the reliability of the essay scores. Specifically, reliabilities were estimated separately for standard essays and for DBQs by asking some students to write three standard essays (two in a special administration and one as part of the national testing program) while other students wrote two DBQs (one in a special administration and one as part of the national administration).

Method

Sample

A random sample of high school teachers with at least 20 students in the college-level Advanced Placement (AP) history courses in United States History or European History were asked to administer specially constructed essay tests in their classrooms within two week of the national AP administration in May. Scores on these experimental essays were then matched with scores from the national administration. After matching, the final sample consisted of 1045 US history students and 891 European history students.

Materials and procedures

Within both subject areas (U.S. history and European history), students were randomly assigned to write in one of six question booklets. Two of these booklets contained a DBQ, with a different DBQ topic in each of the two booklets. Topics for these DBQs are presented in Appendix B. The other four booklets each contained two standard essay topics that were selected to represent different eras and different emphases (e.g., social/intellectual history or political/economic history). Four topics were arranged in four pairs. Order was counterbalanced so that the topic that was listed first in half of the pairs was second in the other half, although with four topics and four pairings not all possible combinations could be represented. Specifically, topic 1 was administered together with topic 3 or topic 4, and topic 2 was administered with topic 3 or topic 4, but topic 1 was not administered with topic 2 and topic 3 was not administered with topic 4.

The testing session lasted one hour. Students with the DBQ booklets were asked to spend 15 minutes reviewing the documents and planning their answers and the remaining 45 minutes

writing. Students with the standard essay booklets were asked to spend about 30 minutes responding to each of the two questions. Students were also asked to indicate which topic they preferred; results for this preference analysis are described in Bridgeman, Morgan, and Wang (in press). The one hour total time limit was enforced, but no check was made of whether the students followed the suggestion for how to allocate their time.

For the regular national administration, the DBQ administration started with a mandatory 15 minute review period followed by 50 minutes (for U.S. History) or 45 minutes (for European History) to write the essay. For the standard essay, U.S. History students were required to choose one of five topics provided and write the essay in 50 minutes; European History students had a choice of one of six topics and 45 minutes to write the essay.

Scoring. Each DBQ essay from the experimental testing was holistically scored on a 15-point scale by the same pool of readers, and at the same time, as the regular national AP scoring. Similarly, each standard essay was read at the national AP scoring, but readers used a 9 point scale for these experimental essays and a 15 point scale for the standard essays from the national administration. (This modification was made in anticipation of a change to shorter essays and a 9 point scale in the following testing year.) Each of the experimental essays in a pair was read by a different reader.

Results and Discussion

Means and standard deviations for the for the two groups that took the U. S. History DBQs in the experimental administration are presented first. For comparison, the means of the three scores from the national test administration are listed separately for the entire population tested nationally as well as for the two experimental samples. The experimental samples had

slightly higher mean scores than the population and comparable standard deviations, and appeared to be well within bounds for making meaningful generalizations. The means for both DBQs administed in the special administration were comparable to the means for the national test, suggesting that students were motivated to perform on the special tests.

Table 2 presents similar information for the two European History DBQ samples. These samples appear to be even more comparable to the population means and standard deviations. And once again, scores from the special administration appear to be at least as high as from the national test.

Correlations of DBQ scores from the special administration with DBQ and standard essay scores from the national administration are presented in Table 3. If the DBQ were assessing a unique skill that went beyond the skills needed to answer standard essay questions, the correlations among DBQs might be expected to be higher than the correlation of the DBQs with the standard essays. However, this was not the case. In both samples, the correlation of the special DBQ with the national DBQ was slightly (but not significantly) lower than its correlation with the standard essay. Additional skills may still be needed to answer the DBQ, but it appears that individual differences in these skills can be predicted as well from performance on a standard essay as from performance on another DBQ.

Correlations for the European History samples are provided in Table 4. The conclusion is the same. There is no evidence that individual differences on the DBQ are more related to differences on another DBQ than they are to differences on a standard essay.

Because DBQ scores were as related to standard essay scores as they were to scores from another DBQ, the correlation of standard essay scores and DBQ scores could suffice as an

indicator of DBQ reliability. However, this would not preclude the possibility that the reliability of standard essay scores could be underestimated by this method. Correlations among standard essay topics and the correlations of those topics with the DBQ score are presented next.

Table 5 presents the means and standard deviations for the U. S. History sample that wrote standard essays in the special administration. Because this group was a random sample from the same classrooms as the DBQ sample, the means and standard deviations on the national test scores should be comparable to those on Table 1 (and they are). Recall that the topics in the special administration were evaluated with a 9 point scale while the standard essays in the national administration were scored on a 15 point scale. Thus, the means on the essays from the special administration were lower. Topic 2 (Civil War) appeared to generate somewhat higher scores, but otherwise the topic difficulty (and/or scoring severity) were comparable.

Table 6 presents comparable information for the European History sample. Although the mean score for the standard essay in the national test was about the same in the U.S. History and European History samples, the means on the essays in the special administration were lower in the European History. Inspection of histograms for the four topics suggested that the top three score categories were rarely used. Either these topics were unusually difficult or (more likely) the raters were not yet adjusted to distributing scores on the 9 point scale.

The correlations are presented in Table 7. Except for the unusually low correlations involving Topic 3 (U.S. as a postwar power), the topics appeared to be fairly comparable in terms of their correlational patterns. This topic may have been an outlier because not all classes had spent any appreciable time on the postwar period at the time of the testing at the end of April. Even though the topic demonstrated reasonable individual differences, they may have been more

related to outside learning than to classroom experiences. Scores from both essays taken by an individual were combinded to form a total essay score for that individual. But because of the anomolous performance of Topic 3, the total was not computed for students who had that topic. Thus, the last line in the table totals the score only for those student who had Topics 1 and 4 or 2 and 4. This total is slightly (but not significantly) more highly correlated with the DBQ score than with the score on the standard essay, suggesting that using the correlation of the standard essay and the DBQ to estimate the reliability of the standard essay does not result in an underestimate of the standard essay's reliability. For this sample of 664 students, the correlation of the standard essay from the national administration with the DBQ from the national administration was .29.

Table 8 presents the comparable correlations for the European History sample. All topics appeared to function comparably in terms of correlational patterns. Even though Topic 3 (postwar economic revival) appeared to be especially difficult (see Table 6), this did not appear to adversely affect the correlations. Although the total for both topics appeared to correlate slightly better with the standard essay (.41) than with the DBQ (.34), this difference was not significant ($t = 1.62$; 95% confidence interval for the difference -.02 to .20). Furthermore, the correlation of the standard essay and DBQ (both from the national test) in this sample was .29 which is quite consistent with a reliability estimate based on the correlation between topics.

Conclusion

Individual differences on a given DBQ are as predictable from scores on a standard essay as they are from scores on a different DBQ. If additional skills are required to answer a DBQ that are not needed to answer a standard essay, it appears that these skills are possessed by students who also do well on the standard essays. Similarly, standard essay scores are correlated about as

highly with the DBQ as with other standard essay scores. The current practice of estimating the reliability of the essays from the correlation of the standard essay and the DBQ does not appear to result in an underestimate. From a broader perspective, these results reinforce the notion that topic effects are more important than mode effects.

References

Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). Assessing writing skill. (Research Monograph No. 11). New York: College Entrance Examination Board.

Bridgeman, B., Morgan, R., & Wang, M. (in press). Choice among essay topics: Impact on performance and validity. Princeton, NJ: Educational Testing Service.

Carlson, S. B., Bridgeman, B., Camp, R., & Waanders, J. (1985). Relationship of admission test scores to writing performance of native and nonnative speakers of English (TOEFL Research Report No. 19 and RR-85-21). Princeton, NJ: Educational Testing Service.

College Board. (1993). Advanced Placement course description: History--United States History, European History. New York: Author.

Quellmalz, E., Capell, F., & Chou, C. (1982). Effects of discourse and response mode on the measurement of writing competence. Journal of Educational Measurement, 19, 241-258.

Appendix A

Standard Essay Topics

U.S. History
Standard Essay Topics

Topic 1. Access the impact of THREE of the following on the status of African Americans from the end of Reconstruction to 1900.

The Fourteenth Amendment
"Black Codes"
Plessy v. Ferguson
The Atlantic Compromise

Topic 2. Identify THREE of the following and evaluate the relative importance of each of the THREE in laying the groundwork for the Civil War.

Abolitionism
The Mexican War
The Kansas-Nebraska Act
The Dred Scott decision

Topic 3. Analyze the ways in which THREE of the following called into question United States preeminence as a global power.

The post-war reconstruction of Germany and Japan
Nuclear proliferation
The Vietnam War
The Organization of Petroleum Exporting Countries (OPEC)

Topic 4. Analyze the relative importance of religious dissent and demographic change in undermining the Puritan dream of establishing a godly and orderly society in seventeenth century New England.

European History
Standard Essay Topics

Topic 1. "In seventeenth-century England the aristocracy lost its privileges but retained its power; in seventeenth-century France the aristocracy retained its privileges but lost its power".

Assess the accuracy of this statement with respect to political events and social developments in the two countries in the seventeenth century.

Topic 2. To what extent and in what ways has twentieth-century physics challenged the Newtonian view of the universe and society?

Topic 3. Assess the strengths and weaknesses of the economic revival of Western Europe between 1945 and 1970.

Topic 4. What were the responses to the Catholic authorities in the sixteenth century to the challenges posed by the Lutheran Reformation?

Appendix B

DBQ Topics

Form UB

Question Booklet

United States History

Document-Based Question

You will have 15 minutes to read the contents of this question booklet. You are advised to spend all of the 15 minutes analyzing the documents and planning your answer for the document-based essay question. You may make notes in this booklet, but only your writing in the answer booklet will be scored. At the end of the 15-minute period you will be told to begin writing on the lined pages of the answer booklet. You will then have 45 minutes to write the essay. Please write your essay with a pen.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

The College Board
Advanced Placement Examinations

UNITED STATES HISTORY

(writing time–45 minutes)

Directions: The following question requires you to construct a coherent essay that integrates your interpretation of Documents A-H and your knowledge of the period referred to in the question. In your essay, you should strive to support your assertions both by citing key pieces of evidence from the documents and by drawing on your knowledge. Write the form code (**UB**) and question number (1) on the top of each page used in the answer booklet.

1. The United States decision to drop an atomic bomb on Hiroshima was a diplomatic measure calculated to intimidate the Soviet Union in the post-Second-World-War era rather than a strictly military measure designed to force Japan's unconditional surrender.

   Evaluate this statement using the documents and your knowledge of the military and diplomatic history of the years 1939 through 1947.

Copyright © 1986 by Educational Testing Service. All rights reserved.

Document B

Source: Memoirs of General H.H. Arnold, Commander of the American Army Air Force in the Second World War (1949)

The surrender of Japan was not entirely the result of the two atomic bombs. We had hit some 60 Japanese cities with our regular H.E. (High Explosive) and incendiary bombs and, as a result of our raids, about 241,000 people had been killed, 313,0000 wounded, and about 2,333,000 homes destroyed. Our B-29's had destroyed most of the Japanese industries and, with the laying of mines, which prevented the arrival of incoming cargoes of critical items, had made it impossible for Japan to carry on a large-scale war....Accordingly, it always appeared to us that, atomic bomb or no atomic bomb, the Japanese were already on the verge of collapse.

Document C

Source: Dwight D. Eisenhower, recollections of a July 1945 meeting with President Harry S Truman (1948)

Another item on which I ventured to advise President Truman involved the Soviet's intention to enter the Japanese war. I told him that since reports indicated the imminence of Japan's collapse, I deprecated the Red Army's engaging in that war. I foresaw certain difficulties arising out of such participation and suggested that, at the very least, we ought not to put ourselves in the position of requesting or begging for Soviet aid. It was my personal opinion that no power on earth could keep the Red Army out of that war unless victory came before they could get in.

Document E

Source: British Prime Minister Winston Churchill's recollections of news received during the Potsdam Conference, July 1945 (1953)

On July 17 world-shaking news had arrived....

The atomic bomb is really...Here then was a speedy end to the Second World War, and perhaps to much else besides...Up to this moment we had shaped our ideas towards an assault upon the homeland of Japan by terrific air bombing and by the invasion of very large armies...

Now all this nightmare picture had vanished. In its place was the vision–fair and bright indeed it seemed–of the end of the whole war in one or two violent shocks...

Moreover, we should not need the Russians. The end of the Japanese war no longer depended upon the pouring in of their armies for the final and perhaps protracted slaughter. We had no need to ask favours of them. A few days later I mentioned to Mr. Eden: "It is quite clear that the United States do not at the present time desire Russian participation in the war against Japan." The array of European problems could therefore be faced on their merits and according to the broad principles of the United Nations. We seemed suddenly to have become possessed of a merciful abridgment of the slaughter in the East and of a far happier prospect in Europe. I have no doubt that these thoughts were present in the minds of my American friends.

Document H

Source: Harry S Truman, radio address (August 1945)

I realize the tragic significance of the atomic bomb.

Its production and its use were not lightly undertaken by this Government. But we knew that our enemies were on the search for it. We know now how close they were to finding it. And we know the disaster which would come to this nation, and to all peaceful nations, to all civilizations, if they had found it first.

That is why we felt compelled to undertake the long and uncertain and costly labor of discovery and production.

We won the race of discovery against the Germans

.

Having found the bomb, we have used it. We have used it against those who attached us without warning at Pearl Harbor, against those who have starved and beaten and executed American prisoners of war, against those who have abandoned the pretense of obeying international laws of warfare. We have used it in order to shorten the agony of war, in order to save the lives of thousands and thousand of young Americans.

We shall continue to use it until we completely destroy Japan's power to make war. Only a Japanese surrender will stop us.

Form AE

Question Booklet

European History

Document-Based Question

You will have 15 minutes to read the contents of this question booklet. You are advised to spend all of the 15 minutes analyzing the documents and planning your answer for the document-based essay question. You may make notes in this booklet, but only your writing in the answer booklet will be scored. At the end of the 15-minute period you will be told to begin writing on the lined pages of the answer booklet. You will then have 45 minutes to write the essay. Please write your essay with a pen.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

The College Board
Advanced Placement Examinations

EUROPEAN HISTORY

(writing time–45 minutes)

Directions: The following question is based on the accompanying Documents 1-13 (Some of the documents have been edited for the purpose of this exercise.) Write your answer on the lined pages of the answer booklet.

Write the form code (**AE**) and the question number (**1**) at the top of each page used in the answer booklet.

This question is designed to test your ability to work with historical documents. As you analyze each document, take into account its source and the point of view of the author. Write an essay on the following topic that integrates the analysis of the documents. You may refer to Historical facts and developments not mentioned in the documents.

1. To what extend did changing views on the causes of juvenile crime affect legal treatment of the juvenile offender in the nineteenth-century Great Britain?

   Historical background: Early in the nineteenth century, the British government, responding to the public's fear that lawlessness was increasing, took measures to reduce crime. For example, it established a professional police force and created committees to study the justice system. Generally, punishments were based on the assumption that crime stemmed from an individual's weak or evil character.

Copyright © 1986 by Educational Testing Service. All rights reserved.

Document 3

At the present time, there are three children, one 7, one 8, and another 10, all of whom were undergoing confinement in separate cells in the Clerkenwell penitentiary. The youngest child has been convicted at Manchester of having stolen certain goods which the mother had received, knowing them to be stolen property. The child was sentenced to transportation and the mother to six months' imprisonment. Since the child's years would furnish sufficient proof that she could not be fully aware of the nature of her offense, the child's sentence was commuted to solitary imprisonment under the separation system.

Testimony of Mr. Hoare, a visiting justice at the Clerkenwell House of Correction, to the London Court Sessions, 1838

Document 6

Young thieves have often confessed to me that their first attempts at stealing began at apple stalls. Acquiring confidence by a few successful adventures, they have gradually progressed in crime. They find companions to cheer them and instruct the, girls to share their booty and applaud them. Imprisonment is no punishment. It's no matter to him where he exists as long as he has food and some clothing. In fact, many lads have admitted to me that they learned more in jail than out of it.

Mr. W.A. Miles, Esquire, *Report on Prison Discipline,*, presented to the House of Lords, 1835

Document 7

One grand cause of depravity and crime in children is the vice of their parents, who often educate their offspring in the art of thieving and live upon the proceeds of their children's depredations. In speaking of three children whose lives of crime he relates, Mr. Rushton observes, "These lads have been trained by a vicious father to the work of plunder. He has taught them how to steal with dexterity, and he uses them as a means of supplying himself with a luxurious existence."

Unless the evil power of the parent be destroyed and his mischievous teaching counteracted, it is clear that no valid hopes of reformation can be obtained.

Report of an address to the Town Council of Liverpool by its magistrate, Mr. Rushton, *The Times* of London, August 21, 1850

Document 8

The scanty wages given to many forms of labor, as well as the high price of rent and provisions, make it almost impossible for a man alone to support the family. Hence, most of the wives of the unskilled workpeople have to forego their maternal duties, and devote themselves to some kind of drudgery to add to the petty household income. If then the mother be away from home the greater part of her time, and the children be left to gambol in the gutter with others as neglected, what reward can society look for from this moral anarchy and destitution? Here is the real explanation of juvenile delinquency.

H. Mayhew and J. Binny, *The Criminal Prisons of London*, London, 1862

Document 12

Industrial Schools are credited, we believe justly, with having broken up the gangs of young criminals in the larger towns,; with putting an end to the training of boys as professional thieves; and with rescuing children fallen into crime from becoming habitual or hardened offenders. Undoubtedly, they have also had the effect of preventing a large number of children from entering a career of crime.

Royal Commission Report,
Parliamentary Papers, 1884

Document 13

The proportion of children now sent to prison is much smaller than it was. Besides this, the sentences passed on children are much lighter than they were, and a check is exercised on the Magistrates. As an instance, I will quote the case of a boy sentenced to a month's imprisonment for stealing fruit. His term of imprisonment was changed from a month to seven days by the Home Office which reviews Magistrates' sentences. The practice of flogging nevertheless still exists, and not long ago a policeman told me, with relish, that he had birched [whipped] as many as sixty boys on a single day at one court session.

Still, there is a growing tendency to make grave distinctions between the treatment of juvenile and adult criminals. An outcome of this tendency may be found in the creation of Reformatories and Industrial Schools for younger and, as a rule, more innocent children. A system of rewards in money obtains both in Reformatory and Industrial Schools, so that by good conduct a child may accumulate a small sum for the time when its period of detention shall be over. The necessity for corporal punishment is minimized by the system of rewards.

Gertrude M. Tuckwell, *The State and Its Children*,
London, 1894